

Utilisation de la norme ISO11179 pour améliorer l'interopérabilité entre les différents schémas de métadonnées pédagogiques

Nicolas Delestre

Laboratoire PSI-INSA de Rouen
BP 08 Place Émile Blondel, 76131 Mont Saint Aignan Cedex, France
Nicolas.Delestre@insa-rouen.fr

Yolaine Bourda

École Supérieure d'Électricité
3 rue Joliot Curie, 91192 Gif-sur-Yvette Cedex, France
Yolaine.Bourda@supelec.fr

Résumé

Depuis Juillet 2002, le LOM est devenu un standard de l'IEEE, mais ce n'est pas le seul schéma de métadonnées pédagogiques existant. En effet, d'autres schémas totalement indépendants sont aussi utilisés, et depuis quelques mois de nombreux schémas issus du LOM apparaissent : on les nomme des profils d'applications du LOM. Or en les étudiant, on en déduit que l'interopérabilité entre ces schémas n'est plus automatique. Pour résoudre ce problème, nous proposons d'utiliser une méthode définie par la norme ISO11179 qui permet de dissocier l'aspect conceptuel de l'aspect représentation lors de la conception de schéma de métadonnées. Ainsi, grâce à ce modèle, les schémas de description de documents pédagogiques pourraient être considérés comme des instanciations d'un seul et unique modèle conceptuel, et donc devenir interopérables.

Mots-clés

Métadonnée, LOM, ISO11179, interopérabilité

Abstract

The IEEE LOM is now standardized since July 2002 but it is not the only one standard used to describe learning objects. Application profiles based on the LOM and other educational metadata schemas coexist. This established fact leads to a lack of interoperability between descriptions based on different schemas. To solve this problem, we propose a solution based on the ISO 11179 standard, allowing the description of a conceptual level independently of any particular representation. Thus, each educational metadata schema can be specified as an "instantiation" of a unique educational conceptual level linking them.

Key words

Metadata, LOM, ISO11179, interoperability

Contexte de travail : Les schémas de métadonnées pédagogiques

Les métadonnées pédagogiques facilitent la recherche, l'échange et la gestion des données pédagogiques comme des cours, des éléments de cours, des exercices, des exemples, des transparents, etc. [BH00]. L'ensemble d'éléments de métadonnées pédagogiques le plus connu, le LOM (pour *Learning Object Metadata*) est standardisé depuis Juillet 2002 par l'IEEE (Cf. [P1402]). Depuis plusieurs mois on voit apparaître de nouveaux schémas basés sur le LOM, mais qui utilisent aussi d'autres schémas de métadonnées, tel que le Dublin Core : ce sont les "profils d'application". Plus précisément, Heery et Patel dans [HP00] définissent les profils d'application comme étant "des métadonnées issues d'un ou plusieurs schémas de métadonnées combinés afin d'améliorer et d'optimiser leur utilisation dans un cadre particulier"¹. De leur côté, Duval et al. dans [DHSW02] considèrent que l'objectif d'un profil d'application est "d'adapter et de combiner des schémas existants afin d'obtenir un nouveau schéma pour une application particulière tout en gardant l'interopérabilité avec le ou les schémas de base"¹.

Dès lors, comme l'indiquent [HP00] et [DHSW02], un profil d'application peut :

- modifier la définition d'un élément du schéma de métadonnées (que l'on nomme aussi élément de donnée, ou DE pour *Data Element*) afin par exemple de le rendre moins spécifique,
- restreindre l'espace de valeurs d'un DE,
- contraindre un DE à devenir obligatoire alors qu'il était optionnel (mais pas l'inverse),
- définir des relations entre DE,

¹Traduction française

- ajouter de nouveaux DE (il doit alors les maintenir).

La plupart des profils d'application existants sont en fait des profils d'application basés sur le LOM (SCORM, CLEO, EdNA, Celebrate, etc.), ils devraient donc tous être compatibles les uns avec les autres tant au niveau des DE utilisés (on pourra trouver dans [FN03] une présentation des DE utilisés par le LOM), qu'au niveau du vocabulaire utilisé pour expliciter chaque DE. Est-ce le cas ? Malheureusement non. Pour illustrer ce problème prenons l'élément de donnée nommé *Learning Resource Type* (DE 5.2 du LOM) :

- dans le LOM cet élément est défini de la façon suivante : “Spécifie le type d'un objet pédagogique. Le type dominant est celui qui doit apparaître en premier”. Les valeurs associées à ce DE sont, en anglais : *exercise, simulation, questionnaire, diagram, figure, graph, index, slide, table, narrative text, exam, experiment, problem statement, self-assessment, lecture*. Ce DE peut donc être utilisé pour décrire le type physique d'un document pédagogique aussi bien que son potentiel pédagogique. De plus le LOM définit ce DE comme étant compatible avec le DE nommé *type* du DCMI [DCM03] alors que les valeurs admises (toujours en anglais) par ce dernier sont : *Collection, Dataset, Event, Image, InteractiveResource, Service, Software, Sound, Text, PhysicalObject, StillImage, MovingImage*.
- dans CLEO [CLE03], ce DE peut prendre les valeurs suivantes : *additional resource, analogy, assessment, assessment item, attractor, community, definition, demonstration, example, feedback, glossary, guidance, guideline, illustration, importance, introduction, non example, note, objective, outline, overview, practice, prerequisite, presentation, recall, reference, reinforcement, scenario, summary*.
- dans RDN/LTSN [RDN03], ce DE peut prendre les valeurs suivantes : *CourseModuleUnitProgramme, ResourcePack, CaseStudy, ActivityExerciseFieldwork, ActivityExerciseFieldworkNotes, ProjectOutline, SimulationModel, StudyGuide, ExaminationTest, QuestionBank, AssessmentItem, LecturePresentation, Glossary, CourseLecturePresentationNotes, WorkedExample, textbook, Demonstration, Computer-basedTutorial, ReadingList, TeachingTip, CurriculumSyllabus, EducatorGuide, EvaluationForm, EducationalPolicy, EducationalReport, LessonPlan*.
- enfin dans le schéma de métadonnées de l'EdNA [EdN00], ce DE est défini comme étant “la nature ou le genre du contenu de la ressource” et il peut prendre les valeurs provenant du vocabulaire DCMI ou des valeurs spécifiques d'EdNA tels que *edna-document, edna-curriculum* ou *edna-event*.

On voit bien à travers cet exemple qu'un même DE,

ou du moins un DE qui porte le même nom, peut prendre des valeurs bien différentes.

De la même manière, pour le CLEO, une illustration est définie comme étant “un graphique ou un autre média de présentation”. Cela correspond-il à la valeur *diagram* pour le LOM, ou à la valeur *figure* ou encore *graph* ?

Pourquoi de telles différences ? Tout simplement parce que ces schémas de métadonnées sont basés sur des termes et non sur des concepts. En effet aucun d'eux ne définit explicitement les concepts présentés par chaque DE. Il est donc pratiquement impossible de pouvoir traduire automatiquement des DE d'un schéma dans un autre.

Comment peut-on résoudre ce problème ? En repensant totalement la conception d'un schéma de métadonnées et en utilisant des méthodes permettant de créer ce dernier de manière non ambiguë : c'est ce que nous proposons de faire à l'aide de la norme ISO11179.

La norme ISO 11179

La première version de la norme ISO11179 [ISO03] date de 1994, et elle est depuis en constante évolution. Cette norme est composée de plusieurs parties :

1. “Cadre de la spécification et de la normalisation des éléments de données” (dernière version 1999) qui définit le cadre d'utilisation de cette norme.
2. “Classification des éléments de données” (dernière version 2000) qui définit comment classer des DE.
3. “Métamodèle de registre et attributs de base” (dernière version 2003) que nous allons expliciter par la suite.
4. “Règles et directives pour la formulation des définitions des données” (dernière version 1995) qui définit des règles de nommage pour les différents éléments d'un schéma de métadonnées.
5. “Principes de dénomination et d'identification des données” (dernière version 1995)
6. “Enregistrement des éléments de données” (dernière version 1997)

L'objectif de cette norme, surtout sa troisième partie, est de définir une méthode de spécification de schéma de métadonnées. Cette méthode repose sur quatre niveaux :

1. Le niveau conceptuel permettant de définir les concepts et leurs caractéristiques issus du domaine que l'on veut représenter.
2. Le niveau métadonnées conceptuelles où l'on identifie les métadonnées de notre domaine de manière conceptuelle, c'est-à-dire en dehors de tout contexte d'utilisation.
3. Le niveau représentation où l'on définit, pour un contexte particulier, comment on représente les métadonnées conceptuelles.

4. Le niveau classification qui permet d'organiser les métadonnées.

Étudions plus en détail ces quatre niveaux à travers un exemple basé sur la construction d'un schéma permettant de représenter les noms de pays.

Le niveau conceptuel

C'est à ce niveau, dont le modèle UML est représenté par la figure 1, que l'on identifie toutes les données qui permettent de décrire les entités du domaine considéré. Ces entités sont des instances de *object class* (plus exactement instances de *Concept* reliées par des instances de *Concept_Relationship*) et les propriétés de ces entités sont des instances de *property*. En fait, les termes *object class* et *property* ont été utilisés en lieu et place des habituels termes classes et attributs pour ne pas confondre les classes et attributs du modèle de la norme (représentés à l'aide du langage UML) et les classes et attributs du domaine que l'on veut représenter (qui vont être des instances du diagramme UML de la norme). La relation avec cette vision informatique est tellement forte que dans les exemples proposés par la littérature effectuant une analyse de haut en bas (*top-down analysis*) comme par exemple [ISN01], la première étape spécifie un modèle conceptuel en utilisant UML comme langage de représentation. Ensuite, à partir de ce modèle UML sont extraits les éléments de ce niveau conceptuel. On verra dans la dernière partie de ce papier que cela nous aidera à construire une chaîne de production de modèles compatibles avec l'ISO 11179 (du moins la partie 3 de cette dernière).

Dans notre exemple, nous pouvons décrire le concept de pays en le reliant à quatre propriétés : identifiant, capitale, langues et nombre d'habitants.

Le niveau métadonnées conceptuelles

Une fois le modèle conceptuel défini, il faut identifier les métadonnées qui permettent de décrire les entités de notre domaine, ainsi que les valeurs conceptuelles que ces dernières pourront prendre. Pourquoi utiliser le terme "conceptuel" ? Car ces valeurs sont en théorie indépendantes de toute représentation, c'est-à-dire de tout contexte d'utilisation, aussi bien culturel qu'applicatif.

Chaque métadonnée conceptuelle (DEC pour *Data Element Concept*) est l'union d'une instance d'*object class* (donc un concept ou une relation entre concepts) et d'une instance de *property* du niveau précédent. Comme l'indique la figure 2, à chaque DEC est associé un domaine de valeurs conceptuelles possibles (CD pour *Conceptual Domain*), ce CD pouvant regrouper des valeurs énumérées (plusieurs instances de *Value_Meaning*) ou pouvant être numérique (au sens large : nombre, date, intervalle, etc.).

Dans notre exemple, nous avons quatre DEC :

1. Le DEC "identifiant de pays" dont le domaine conceptuel est énuméré (pour les 230 pays).
2. Le DEC "capitale de pays" dont le domaine conceptuel est énuméré.
3. Le DEC "langues de pays" dont le domaine conceptuel est énuméré.
4. Le DEC "nombre d'habitants d'un pays" dont le domaine conceptuel est non énuméré et de type entier.

Le niveau représentation

Le niveau représentation décrit quant à lui une utilisation particulière (dans un contexte donné) du modèle précédent (Cf. figure 3). Ainsi tout ou partie des DEC sont projetés vers des *Data Element* (DE) en explicitant les valeurs que peuvent prendre ces derniers, celles-ci constituent les *Value Domain* (VD). Ces VD sont des projections des CD du niveau précédent.

Ainsi dans notre exemple, on peut considérer que la norme 3166 (Cf. [ISO04]) est constituée de plusieurs représentations de notre modèle conceptuel. En effet, la norme ISO3166 propose d'identifier les pays sous différentes formes : les noms des pays en anglais (version courte ou étendue), les noms des pays en code anglais (2 ou 3 lettres), les noms des pays en français (version courte ou étendue) et les noms des pays en code français.

Le niveau classification

Une fois ces trois premiers niveaux explicités, la norme propose d'organiser ces informations en les associant à une ou plusieurs classifications qui peuvent être hiérarchique, ontologique, etc.

L'ISO 11179 et les schémas de métadonnées pédagogiques

Comme nous l'avons vu dans la première partie de ce papier, l'interopérabilité entre les schémas de métadonnées est un problème important. Que signifie le mot interopérabilité ? Tout simplement le fait qu'une entité pédagogique référencée dans un système puisse être réutilisée par un autre système possédant un schéma de métadonnées différents du premier. Ceci oblige à l'heure actuelle à posséder un traducteur qui est capable de transformer les métadonnées issues du premier schéma en métadonnées pour le deuxième schéma. Par extension, cela signifie, que s'il existe n schémas de métadonnées, il faut posséder $n * (n - 1)$ traducteurs pour pouvoir être totalement interopérable : ce n'est pas raisonnable. De plus cela sous-entend l'hypothèse que toute métadonnée d'un schéma de métadonnées a son équivalent dans un autre schéma (équivalent direct ou calculé).

A contrario, si l'on utilise la norme ISO 11179, cela implique que nous avons deux niveaux : un niveau conceptuel et un niveau de représentation. En supposant que

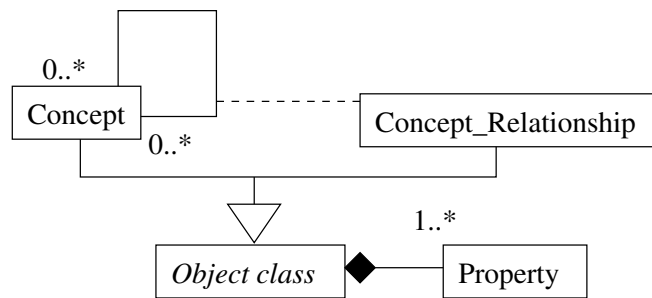


FIG. 1 – Diagramme de classes UML du niveau conceptuel de l'ISO11179-3

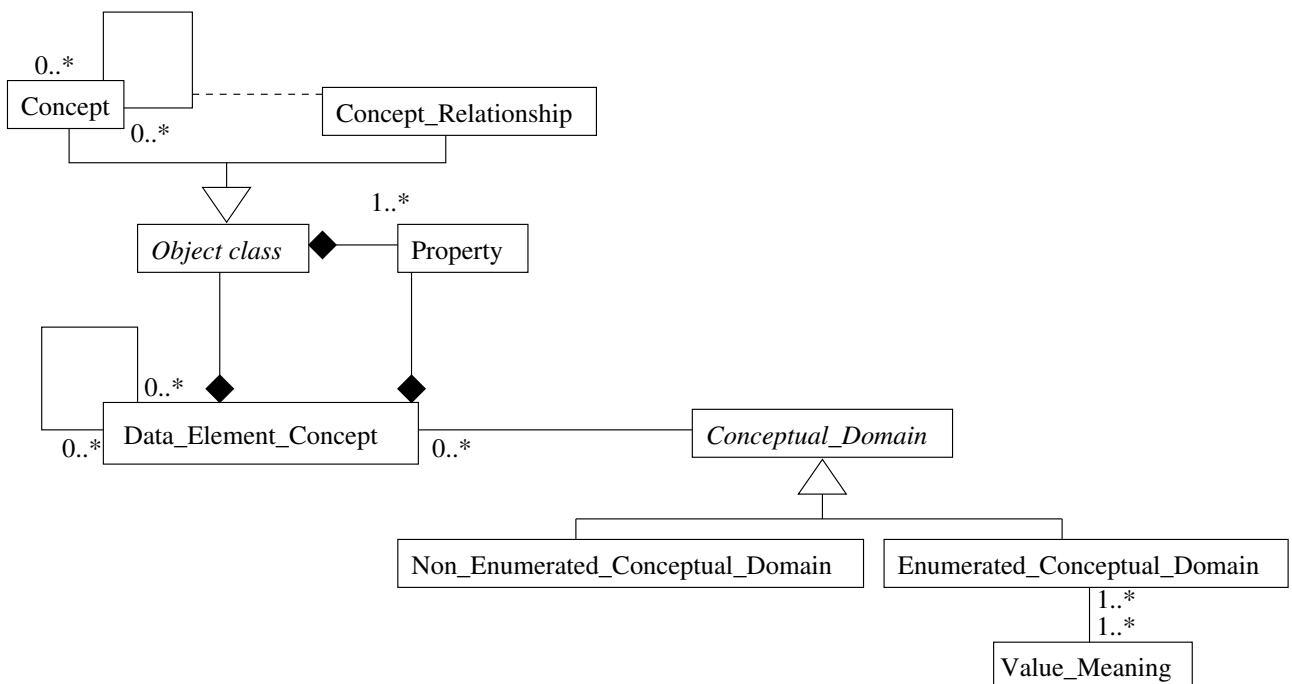


FIG. 2 – Diagramme de classes UML du niveau metadonnées conceptuelles de l'ISO11179-3

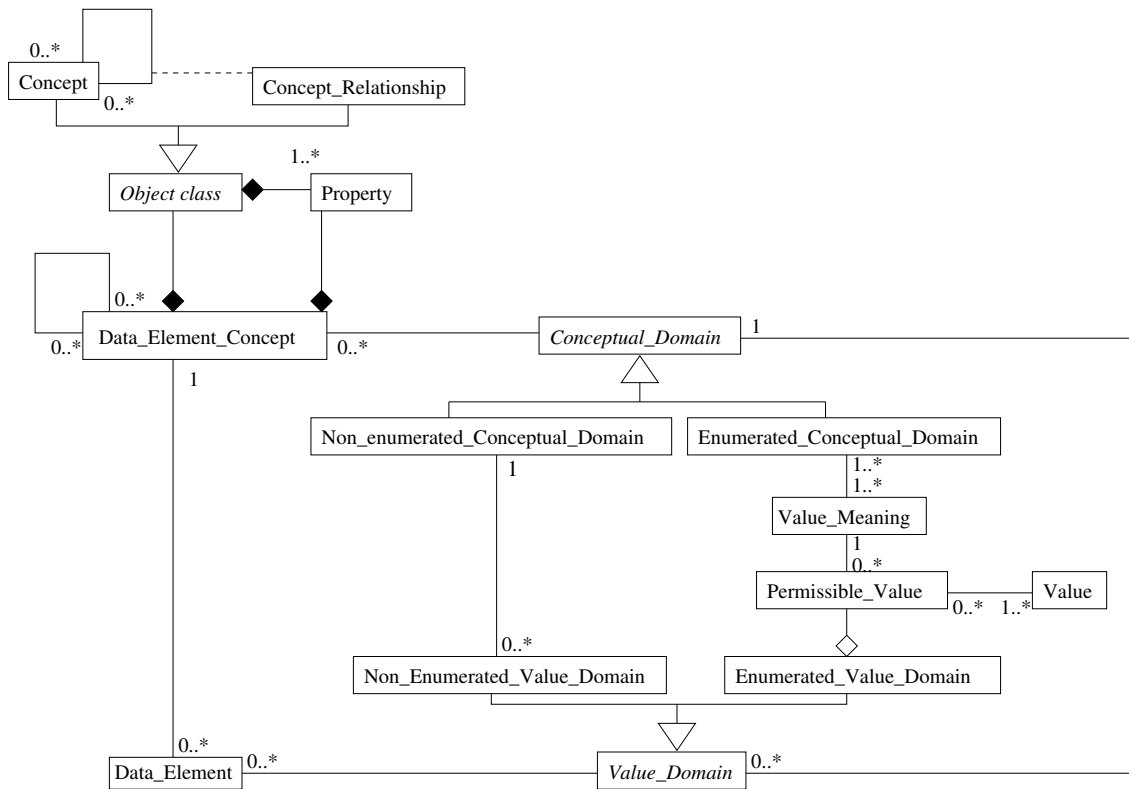


FIG. 3 – Diagramme de classes UML de l'ISO 11179-3

l'on soit capable de décrire de manière conceptuelle les métadonnées qui caractérisent un document pédagogique, on peut considérer que les schémas de métadonnées actuels sont des représentations de ce schéma de métadonnées conceptuelles. Cela implique alors que pour n systèmes d'enseignement possédant chacun son schéma de métadonnées, il faudrait $2n$ traducteurs : n pour traduire les métadonnées des schémas vers le schéma conceptuel et n autres pour les traductions inverses.

C'est vers cette solution que tend le groupe d'experts GE4 de la CN36 de l'Afnor, en nommant ce schéma de métadonnées conceptuelles le MLR pour *Metadata for Learning Ressources*. À l'heure actuelle une première version du MLR est en chantier, où nous mettons en œuvre une analyse mixte : ascendante et descendante. En effet des schémas de métadonnées existent et nous ne pouvons pas en faire fi : il faut absolument que ces derniers soient compatibles avec le MLR. Par conséquent, nous effectuons à la fois une analyse descendante (*top-down*) afin d'être le plus exhaustif possible, et une analyse ascendante (*bottom-up*) afin de prendre en compte les particularités des schémas existants. Par exemple, le DEC 5.2 du LOM, vu dans la première partie de ce document, est présent dans le MLR à travers trois DEC, deux issus de l'analyse descendante (un DEC pour représenter le type physique des documents et un autre pour représenter le type péda-

gogique) et un troisième issu de l'analyse ascendante (qui est l'union des deux premiers d'un point de vue DEC et CD) qui permet au LOM d'être compatible avec le MLR.

Un outil de production

Afin d'aider le développement de schémas de métadonnées compatibles avec la norme ISO11179-3, nous avons modélisé et commencé à implanter un outil basé sur une utilisation des technologies XML. En effet, produire un schéma compatible avec la norme en suivant une analyse *top-down* peut être, comme nous l'avons vu, décomposé en 4 étapes :

1. modélisation conceptuelle du monde que l'on veut représenter
2. identification des *object class* et *property* associés
3. création des DEC avec définition des CD correspondants (des énumérés ou des non-énumérés)
4. enfin, pour un contexte donné, identification des DE et VD associés

On voit alors apparaître que les entités d'un niveau x peuvent être déduites de celles du niveau $x - 1$. D'un point de vue informatique, on peut même dire que les entités d'un niveau x sont des instances de classes identifiées à un niveau $x - 1$ (elles-mêmes instances des classes issues du niveau $x - 2$). Or la création de schémas revient à produire

un ensemble de documents comme le montre par exemple [ISN01], cela revient donc à dire que la grammaire d'un document pour un niveau x est d'une part, issue du document défini au niveau $x - 1$, et d'autre part cadrée par le modèle défini par la norme.

C'est à partir de ces constats que nous avons décidé d'utiliser les technologies XML. En effet, les schémas XML (XSD) sont un modèle de création de grammaire pour les documents XML. Ce modèle a la particularité d'être lui-même défini à l'aide du langage XML et de proposer des mécanismes de modélisation de haut niveau (modularité, héritage, typage fort, etc.). Ainsi les grammaires définies à un niveau x peuvent être issues d'un document XML défini à un niveau $x - 1$ tout en utilisant des entités (définies par la norme) grâce au concept d'héritage. Cette génération, indépendante du schéma que l'on est en train de définir, peut être obtenue à l'aide du langage de transformation XSL-T, défini lui aussi à l'aide du langage XML. Cette dernière caractéristique permet, lorsque le besoin s'en fait sentir, de pouvoir générer automatiquement des feuilles de transformation XSL-T à partir d'autres feuilles de transformation XSL-T.

Ainsi, l'architecture de notre outil de production de schéma de métadonnées compatible avec l'ISO11179-3 est présenté par la figure 4.

On commence par établir le diagramme de classes décrivant le domaine que l'on veut représenter à l'aide des logiciels ad-hoc (comme par exemple ArgoUML, <http://argouml.tigris.org/>). On obtient alors un document XML correspondant à ce modèle (plusieurs formats sont disponibles, par exemple UXF ou XMI).

À partir de ce fichier, par transformation XSL-T, on obtient un fichier XML décrivant l'ensemble des *object class* et *property* disponibles. La validation de ce fichier XML est alors contrôlée par un fichier XSD directement issu de la norme. Si on le désire, à ce niveau, on peut générer de la documentation au format PDF à l'aide de XSL-FO (ou autre par exemple XHTML, L^AT_EX, DocBook, etc.).

Ce même principe est répété au niveau supérieur pour la production des DEC et CD : un fichier XML, contenant par défaut l'ensemble des DEC issus des *object class* et *property*, est généré. Il suffit de le compléter en :

- ajoutant des DEC qui sont issus de l'analyse *bottom-up*
- spécifiant les CD.

Au dernier niveau ce n'est pas un fichier XML qui est généré mais une grammaire cadrant la création de schémas de métadonnées. En effet, à ce niveau, le niveau représentation, le concepteur du schéma n'est pas obligé de représenter l'ensemble des DEC en DE, mais les seuls DE qu'il puisse créer doivent être issus des DEC spécifiés au niveau précédent. Du point de vue XML, les seules balises utilisables sont issues des DEC comme le montrent

les deux extraits de code XML suivants qui reprennent l'exemple présenté dans la deuxième partie de ce document.

DEC et CD

```
...
<Data_Element_Concept>
  <administered_item_administration_record
    registration_status="Standard"
    administrative_status="Final">
    <administered_item_identifier
      data_identifier="Country" version="">
      <item_registration_authority_identifier
        international_code_designator=""
        organization_identifier="">
    ...
  <Enumerated_Conceptual_Domain>
  ...
  <Value_Meaning
    value_meaning_identifier="1001"
    value_meaning_description="The primary geopolitical
      entity known as Afghanistan"
    value_meaning_begin_date="1997-01-10"/>
  <Value_Meaning value_meaning_identifier="1002"
  ...
  </Enumerated_Conceptual_Domain>
</Data_Element_Concept>
```

DE et VD

```
...
<Data_Element_Country>
  <administered_item_administration_record
    registration_status="Standard"
    administrative_status="Final">
    <administered_item_identifier
      data_identifier="EnglishShortName"
      version="">
    <item_registration_authority_identifier
      international_code_designator=""
      organization_identifier="">
    </administered_item_identifier>
  </administered_item_administration_record>
  <Enumerated_Value_Domain_Country
    value_domain_maximum_character_quantity="44"
    value_domain_format="A(44)">
  ...
  <value_domain_datatype
    datatype_name="Alphanumeric"
    datatype_scheme_reference="">
  <value_domain_unit_of_measure
    unit_of_measure_name=""
    unit_of_measure_precision="1"/>
  <Permissible_Value_1001
    permissible_value_begin_date="1997-01-10">
    Afghanistan
  </Permissible_Value_1001>
  <Permissible_Value_1002
    permissible_value_begin_date="1997-01-10">
    Albania
  </Permissible_Value_1002>
  ...
  </Enumerated_Value_Domain_Country>
</Data_Element_Country>
</ISO11179-3>
```

Cet outil, qui est encore un prototype, permet d'une part la création de schémas de métadonnées compatibles avec l'ISO11179-3, puisque le concepteur du dit schéma est guidé par une grammaire, et d'autre part de générer automatiquement les documents décrivant ce schéma dans différents formats (HTML, PDF, etc.).

Conclusions

La standardisation du LOM par l'IEEE ne supprime pas les problèmes d'interopérabilité entre les différents schémas de métadonnées. Certains profils d'application du LOM mettent sous le même nom (le même élément de métadonnées) des concepts différents. Parfois, les concepts sont les mêmes mais les vocabulaires sont différents. La solution que nous proposons est basée sur la

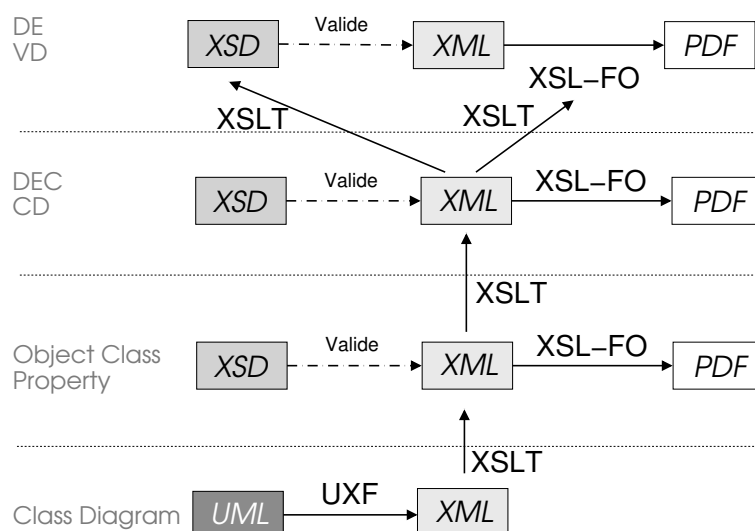


FIG. 4 – Architecture de l’outil de production de schéma compatible avec l’ISO11179-3

norme ISO11179 et sépare l’aspect conceptuel de l’aspect représentation. Ainsi le pivot central permettant une réelle interopérabilité n’est plus le LOM mais un modèle conceptuel indépendant de toute représentation et le LOM devient une représentation comme une autre. Ce modèle conceptuel est en cours de spécification par le groupe GE4 de la CN36 de l’Afnor, et un prototype d’outil d’aide à la conception de schéma de métadonnées compatibles avec l’ISO11179 a été proposé.

Remerciements. Nous remercions deux étudiants du département ASI de l’INSA de Rouen, Céline Capron et Laurent Fallet, pour l’aide qu’ils nous ont apporté pour la conception et le développement de l’outil de production. Nous remercions également les membres du GE4 de la CN36 de l’AFNOR.

Références

- [BH00] Y. Bourda and M. Hélier. What metadata and xml can do for learning objects. *WebNet Journal*, 2(1) :24–31, 2000.
- [CLE03] Extensions to the ieee learning object metadata. http://www.cleolab.org/CLEO_LOM_Ext_v1d1a.pdf, 2003.
- [DCM03] Dublin core metadata initiative type vocabulary. <http://dublincore.org/documents/dcmi-type-vocabulary/>, 2003.
- [DHSW02] E. Duval, W. Hodgins, S. Sutton, and S.L. Weibel. Metadata principles and practicalities. D-Lib Magazine, <http://www.dlib.org/dlib/april02/weibel/04weibel.html>, 2002.
- [EdN00] Education network australia. <http://standards.edna.edu.au/metadata/>, 2000.
- [FN03] N Friesen and L. Nirhamo. Survey of lom implementations. Preliminary Report. ISO/IEC JTC1 SC36 WG4 N0057, 2003.
- [HP00] R. Heery and M. Patel. Application profiles : mixing and matching metadata schemas. Ariadne issue 25, 2000.
- [ISN01] ISO-SC32-N648. Information technology – data management and interchange – part 1. <http://www.jtc1sc32.org/sc32/jtc1sc32.nsf/Attachments/>, 2001.
- [ISO03] ISO11179. Information technology - metadata registries (mdr), 2003.
- [ISO04] ISO3166. Iso 3166 country codes. <http://www.iso.ch/iso/en/prods-services/iso3166ma/index.html>, 2004.
- [P1402] IEEE P1484.12.2/D1. Draft standard for learning technology - learning object metadata. <http://ltsc.ieee.org/wg12/index.html>, 2002.
- [RDN03] Lom application profile. version 1.0. <http://www.rdn.ac.uk/publications/rdn-ltsn/>, 2003.